



Working Paper 06-65
Business Economics Series 19
November 2006

Departamento de Economía de la Empresa
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34) 91 624 9608

YET ANOTHER PUZZLE? THE RELATION BETWEEN PRICE AND PERFORMANCE IN THE MUTUAL FUND INDUSTRY*

Javier Gil-Bazo¹ and Pablo Ruiz-Verdú²

Abstract

Gruber (1996) drew attention to the puzzle that investors buy actively-managed funds even though, on average, they underperform index funds. We uncover another puzzling fact about the market for actively-managed equity mutual funds: funds with worse before-fee performance charge higher fees. We then conduct a series of robustness checks and find that the apparently anomalous fee-performance relation survives all of them. Finally, we show that this relation may be explained as the outcome of strategic fee setting by mutual funds in the presence of investors with different degrees of sensitivity to performance.

Keywords: Mutual fund performance; mutual fund fees; investors' performance sensitivity.

¹ Universidad Carlos III de Madrid, Department of Business Administration. Calle Madrid, 126, 28903 Getafe (Madrid). Spain. E-mail: javier.gil.bazo@uc3m.es

² Universidad Carlos III de Madrid, Department of Business Administration. Calle Madrid, 126, 28903 Getafe (Madrid). Spain. E-mail: pablo.ruiz@uc3m.es

* The authors thank Clemens Sialm, Mikel Tapia, and Ralph Koijen, as well as seminar participants at Tilburg University, Universidad de Navarra, and Universidad Carlos III de Madrid for helpful comments and suggestions. Javier Gil-Bazo also thanks the Finance Department of Tilburg University where part of the paper was written. The usual disclaimer applies. The financial support of Spain's Ministry of Education and Science (SEJ2005-06655/ECON and SEJ2004-01688) is gratefully acknowledged.

A large number of studies have attempted to determine whether managers of equity mutual funds are able to consistently earn positive risk-adjusted returns. Although the ability of at least some managers to earn abnormal returns is still debated,¹ these studies have documented significant differences in risk-adjusted returns across funds. It became apparent early on (Sharpe, 1966), however, that those differences are to a large extent attributable to differences in fund expenses:² fund returns are reported net of expenses, and differences in expenses explain most of the variation in after-expense performance (Carhart, 1997).

Even though the well-documented ability of fees to explain cross-sectional differences in after-fee performance lends support to the hypothesis of an efficient stock market, it also implies that the mutual fund market is informationally inefficient. Somewhat surprisingly, however, most research has been aimed at analyzing whether the remaining cross-sectional variation in performance can be explained by the existence of managers with superior stock-picking skills (see, for instance, Chevalier and Ellison, 1999), while very little effort has been devoted to understanding the fee-performance relation. Given the key role played by the mutual fund market within the financial system,³ investigating the efficiency of the price mechanism in this market is of paramount importance. In this paper, we undertake this task by exploring the relation between fees and before-fee performance in the equity mutual fund industry.

In a well-functioning market, fees would adjust to ensure that, in equilibrium, after-fee performance is equalized across funds. Therefore, in equilibrium, differences in fees would equal differences in before-fee performance, so the slope of a regression of before-fee performance on fees would be one. If fees adjusted only partially to differences in performance, that slope would be positive but less than one. In contrast to this prediction, we find a puzzling negative relation between before-fee performance and fees in a sample of diversified U.S. equity mutual funds: funds with worse before-expense performance charge higher expenses. In an oft-cited article, Gruber (1996) drew attention to the puzzle that investors buy actively managed funds even though, on average, they provide lower after-fee risk-adjusted returns than index funds. Our results uncover yet another puzzling fact about the industry of actively managed mutual funds.

There are several reasons, however, why our initial estimate may not reflect the true relation between before-fee performance and fees. First, our dataset includes both actively managed funds and index funds. Since it is well known that index funds are cheaper than actively managed funds

and that, on average, the former outperform the latter, our results could be due to the presence of index funds in the sample. The puzzle of a negative relation between before-fee performance and expenses would thus reduce to the one identified by Gruber (1996). A similar problem may arise because the dataset contains both funds sold to individual investors and funds that are sold only to institutional investors. Second, the estimated negative relation may be due to a mismeasurement of the fees effectively paid by investors. In our initial estimation—as in most of the work on fund performance—we implicitly assume that expenses are the only fees paid by investors. Taking into account other fees that are often paid by investors could eliminate the negative relation if those fees tend to be lower in funds that charge higher expenses. A third problem is that the sign of the coefficient could be determined by the influence of expensive underperforming funds, which manage just a small fraction of investors’ money and may be short-lived. Finally, the result may be explained by differences between subsectors within the market for actively managed mutual funds. If funds with different investment objectives are not regarded as substitutes by investors, our results could be consistent with a positive relation between fees and before-fee performance within subsectors. Controlling for all these potential problems, however, we find that the negative relation between before-fee performance and fees persists. We then set out to explain this anomalous relation by investigating the role of funds’ performance in the determination of fund fees.

We consider two related hypotheses, which assume that investors differ in their sensitivity to performance. According to the first hypothesis, advanced by Christoffersen and Musto (2002) in the context of money market mutual funds, mutual fund managers set fees taking into account the elasticity of the demand for their shares, so that funds facing less elastic investors charge higher fees. Christoffersen and Musto (2002) argue that funds with worse past performance will face a less elastic demand, since the performance-sensitive investors would have left the fund following bad past performance. If performance is persistent for at least the worse-performing funds (as indicated by Carhart, 1997), this could explain our finding of a negative relation between fees and before-fee performance. An alternative hypothesis, proposed recently by Gil-Bazo and Ruiz-Verdú (2005), is that fund managers with different abilities target different segments of investors. These authors argue that competition among high-ability managers for the money of sophisticated (performance-sensitive) investors will push their fees down and drive the worse-performing funds out of that segment of the market. The latter funds will then target unsophisticated investors,

to whom they are able to charge higher fees. According to this explanation, the reason why underperforming funds charge higher fees is not that their shares are held by unsophisticated investors. Rather, underperforming funds are avoided by sophisticated investors because of their high fees, so they end up in the hands of unsophisticated investors. We test these two hypotheses against an alternative cost-based explanation. According to this explanation, fund characteristics not included in the univariate regression might be associated with both lower management costs and better performance. For instance, if fund size or age are positively correlated with performance and they allow funds to charge lower fees because of scale or learning economies (see e.g., Malhotra and McLeod, 1997), then the negative relation between performance and fees could simply be due to the omission of these variables.

We test these hypotheses in two steps. Building on previous work on fund flows (Sirri and Tufano, 1998; Jain and Wu, 2000; Nanda et al., 2005; Huang et al., 2006), in the first step, we estimate a flow equation that relates fund flows to different fund characteristics. This allows us to obtain for each fund an estimate of the sensitivity of its flows to performance, which we can then include in an equation of mutual fund fee determination. In the second step, we regress fees on funds' performance, flow-performance sensitivity and a number of variables—including size and age—that have been previously identified as determinants of funds' operating costs. Our results support the hypotheses of Chistoffersen and Musto (2002) and Gil-Bazo and Ruiz-Verdú (2005): funds faced with less sensitive investors charge higher fees, yet, even after controlling for performance-sensitivity, funds with lower expected performance set higher fees.

Our results carry important implications, both for individual investors, who are once again reminded of the importance of carefully considering fund fees when making their investment decisions, and, especially, for regulators. First, our results suggest that a significant fraction of investors responds at best sluggishly to differences in after-fee performance. Second, a significant number of funds exploit that fact and charge high fees. Finally, competition in the market for mutual funds, while disciplining those funds who target sophisticated investors, has not been able to prevent funds that cater to performance-insensitive investors from setting high fees nor to quickly drive them out of the market.

The article is organized as follows. Section I characterizes the equilibrium in a well-functioning mutual fund market; Section II describes the dataset and the different fees charged by mutual

funds; Section III explains how we estimate fund performance; Section IV estimates the relation between before-fee performance and fees and performs a number of tests to evaluate the robustness of the results; Section V discusses several explanations for the estimated relation between fees and performance and tests them; finally, Section VI concludes.

I. Mutual Fund Market Equilibrium

In this section, we derive the market equilibrium condition for a frictionless mutual fund market and obtain an estimating equation to test that equilibrium condition.

In a frictionless mutual fund market, equilibrium can be derived using a standard arbitrage argument. Suppose that asset returns in excess of the risk-free interest rate follow a K -factor model and let R_t^F denote the vector of excess returns at time t of the corresponding K factor portfolios. Then, the Arbitrage Pricing Theory (APT) of Ross (1976) states that, for no arbitrage opportunities to exist, the returns in excess of the risk-free rate of any asset j must equal:

$$r_{jt} = \beta_j R_t^F + v_{jt}, \quad (1)$$

where β_j is the vector of asset j 's exposures to the factors (factor loadings or *betas*) and v_{jt} is a zero-mean error term capturing idiosyncratic risk.

If we let α_{it} denote the ability of fund i 's manager to generate before-fee returns above those earned by any portfolio with identical exposure to the risk factors, then fund i 's before-fee return in excess of the risk-free rate is given by:

$$r_{it} = \alpha_{it} + \beta_i R_t^F + v_{it} \quad (2)$$

Finally, defining $\alpha_{it}^n \equiv \alpha_{it} - f_{it}$ as fund i 's net (or after-fee) *alpha*, fund i 's after-fee return in excess of the risk-free rate can be expressed as:

$$n_{it} = (\alpha_{it} - f_{it}) + \beta_i R_t^F + v_{it} = \alpha_{it}^n + \beta_i R_t^F + v_{it} \quad (3)$$

An arbitrage argument then implies that if funds' factor loadings and alphas are known, in equilibrium all funds must have a zero after-fee alpha. To see this, suppose that there existed funds

with positive after-fee alpha ($\alpha_{it}^n > 0$) in a number sufficient to construct a diversified portfolio (if we consider diversified mutual funds, it should not take a very large number of funds to diversify the residual risk). Let α_{pt}^n denote the after-fee alpha of this portfolio and β_p its vector of factor loadings. Then, it would be possible to construct a zero-cost strategy by investing h dollars in portfolio p and selling short h dollars of a portfolio containing the factor portfolios and the risk-free asset with weights equal to β_p and $1 - \beta_p \iota$, respectively, where ι is a $K \times 1$ vector of ones. The expected payoff of this strategy would equal $h\alpha_{pt}^n$. Since α_{pt}^n is strictly positive, such strategy, would approximate an arbitrage as the residual risk of the fund portfolio approaches zero. It follows that there will always be excess demand for shares of mutual funds with positive after-fee alpha. Since mutual fund shares cannot be sold short, a negative alpha would not constitute an arbitrage opportunity. Investors, however, would avoid mutual funds with negative after-fee alpha, since they would be better off investing in a diversified portfolio with the same factor loadings. Therefore, market equilibrium in the market for mutual funds requires that fees adjust to make all after-fee alphas equal to zero⁴ ($\alpha_{it}^n = \alpha_{it} - f_{it} = 0$), or, in terms of before-fee risk-adjusted returns:

$$\alpha_{it} = f_{it} \text{ for all } i. \quad (4)$$

Therefore, in the absence of market frictions, equilibrium requires before-fee alphas and fees to be positively and linearly related. Further, the slope of the linear relation has to be one.

In the presence of market frictions, such as short-selling or borrowing constraints, trading costs, or costly search, there might be small and transitory deviations from condition (4), with some funds offering small and negative after-fee alphas and others offering small and positive alphas. As long as these deviations are not correlated with fund fees, before-fee performance and fees will be, as in equation (4), linearly related and with a unitary slope.

Our first goal is to evaluate whether the relation between fees and before-fee performance approximates the one-to-one equilibrium relation derived above. We estimate the following equation to test our equilibrium condition:

$$\hat{\alpha}_{it} = \delta_{0t} + \delta_1 f_{it} + \epsilon_{it}, \quad (5)$$

where $\hat{\alpha}_{it}$ is our estimate of α_{it} . In Section III, we discuss in greater detail how we estimate alpha. Here, we only note that as long as the measurement error in $\hat{\alpha}_{it}$ is uncorrelated with fees, it will

not introduce any bias in the estimation.

II. Data

A. Mutual Fund Fee Structure

Fund management fees are typically computed as a fixed percentage of the value of assets under management.⁵ These fees, together with other operating costs—such as custody and administrative fees—constitute the so-called fund’s *expenses*, which are deducted on a daily basis from the fund’s net assets by the managing company. When expressed as a percentage of assets under management, these expenses are known as the fund’s *expense ratio*. Thus, the expense ratio closely approximates the notion of fund fee employed in Section I. Fees paid by fund management companies to brokers in the course of the fund’s trading activity are detracted from the fund’s assets, but are not included in the fund’s expense ratio.

On top of the expenses described above, fund investors are often charged one-time fees known as *loads*, which are used to pay distributors. These loads are paid at the time of purchasing (*sales charge on purchases* or *front-end load*) or redeeming fund shares (*deferred sales charge* or *back-end load*) and are computed as a fraction of the amount invested.⁶ Although loads do not pay for fund management services, they do contribute to the cost of acquiring fund shares. It is worth noting that funds often waive at least a fraction of the loads. Therefore, the loads reported in the CRSP database may often be higher than the ones actually paid by investors.

Further, since the 1980s, many funds charge so-called *12b-1 fees*, which, like loads, are used to pay for marketing and distribution costs, but, unlike loads, are not one-time fees, but, rather, are included in the fund’s expense ratio. Since the 1990s, many funds have been offering multiple share classes with different combinations of loads and 12b-1 fees. In particular, class A shares are characterized by high front-end loads and low annual 12b-1 fees, while classes B and C typically have no front-end loads but higher 12b-1 fees and a contingent deferred sales load, which decreases over time. In the case of C shares, back-end loads only apply the first year, while for class B shares back-end loads are reduced at a 1% annual rate. Class B shares are normally converted into class A shares after a period of 6 to 8 years.

B. Description of the Sample

We obtained our data from the CRSP Survivor-Bias Free US Mutual Fund Database for the period December 1961-December 2003 (see Carhart, 1997; Carhart et. al., 2002; and Elton et al., 2001, for detailed discussions of the dataset). The initial sample contained all open-ended mutual funds alive in the 1961-2003 period. From this initial sample, we excluded all funds that could not be confidently described as diversified, domestic equity mutual funds. Thus, we removed money market funds, bond and income funds, and specialty mutual funds, such as sector or international funds. Although classifications vary throughout the period, the resulting funds can be broadly described as growth or growth and income funds.

To obtain our sample of diversified domestic equity mutual funds, we used the information on funds' investment objectives available in the CRSP database. Unfortunately, this information is not consistent throughout the 1961-2003 period. To address this problem, we combined all the information available on funds' investment objectives to create a homogeneous sample for the years 1961-2003 (see the appendix for details). Some of our results, however, are derived only for the years 1992-2003, for which the information on funds' investment objectives is precise and consistent.

From the sample of diversified equity mutual funds, we deleted observations with no information on returns or expenses or with zero expenses. Inspection of the remaining sample showed that there existed observations with values for expenses or returns that were either obvious errors or values that could not have been generated by diversified equity mutual funds. For example, there were observations reporting monthly returns of more than 300% or expenses of more than 40%. Given the large size of the dataset, we searched for these outliers using Hadi's (1994) outlier detection method.⁷

Table I contains summary statistics for our final sample of 538,813 fund-month observations. The mean expense ratio for the whole sample is 1.37 percentage points, with a standard deviation of 0.61. Figure 1 displays the time-series variation of average fees during the sample period, and reveals two facts. First, expense ratios have smoothly grown throughout the sample period (with average expenses increasing from 0.78 in 1962 to 1.5 in 2003), with growth in the late 1980s and 1990s attributable to the introduction of 12b-1 fees. Second, average loads, despite changing little for the first twenty years of the sample, experienced a significant drop in the 1980s, levelling off

by the mid-1990s. To assess the cross-sectional variation in fees, we have also computed standard deviations and coefficients of variation by year (not reported). For expenses, these coefficients of variation average 0.42 over the whole sample with little variation over time. The dispersion in loads (measured by the coefficient of variation), in contrast, has increased over time, even if we restrict attention to funds charging positive loads.

It is interesting to evaluate Figure 1 in the light of Figures 2 and 3. These figures display the dramatic growth of the market for equity mutual funds both as measured by total net assets and by the number of funds. While in year 1962, there were 110 diversified equity mutual funds in our sample, this number had grown to 671 by 1990 and to 5,613 (2,295 if all share classes of a given fund—coded by CRSP as different funds—are counted as one fund) by 2003.⁸ Therefore, it seems that, although the large growth in the number of funds may have led to a reduction in loads, it has not reduced fund expenses.

III. Mutual Fund Performance Estimation

To estimate the equilibrium equation (5), we first estimate fund performance. Following a long list of studies in the mutual fund performance evaluation literature, we employ Carhart’s (1997) model to measure risk-adjusted mutual fund returns.⁹ In order to evaluate the robustness of our results, we also use Fama and French (1993) three-factor model. In either case, we follow Carhart’s (1997) two-stage estimating procedure.¹⁰ In the first stage, we estimate every month each fund’s exposure to risk factors (*betas*) over the previous five years. If less than five years of previous data are available for a specific fund-month, we require that the fund has been active for at least 48 months in the previous five years, and then use the available data to estimate its betas. In particular, factor exposures are estimated as the slope coefficients in the OLS regressions:

$$r_{is} = \beta_{0,it}^{FF} + \beta_{rm,it}^{FF}rm_s + \beta_{smb,it}^{FF}smb_s + \beta_{hml,it}^{FF}hml_s + \varepsilon_{is}^{FF} \quad (FF)$$

$$r_{is} = \beta_{0,it}^C + \beta_{rm,it}^Crm_s + \beta_{smb,it}^Csmb_s + \beta_{hml,it}^Chml_s + \beta_{pr1y,it}^Cpr1y_s + \varepsilon_{is}^C, \quad (C)$$

where the first equation estimates factor exposures according to Fama and French (1993) three-factor model, and the second one estimates factor exposures according to Carhart’s (1997) model.

In both expressions, r_{is} is fund i 's before-expense return¹¹ in month s ($s = t - 60, t - 59, \dots, t - 1$) in excess of the risk-free interest rate proxied by the 3-month T-Bill secondary market rate; rm_s is the market portfolio return in excess of the risk-free rate; and smb_s , hml_s , $pr1y_s$, denote the return on portfolios which proxy for common risk factors associated with size, book-to-market and momentum effects.¹²

In the second stage, we estimate performance as the difference between before-expense returns and model-implied returns given the fund's estimated exposure to risk factors:

$$\hat{\alpha}_{it}^{FF} \equiv r_{it} - \hat{\beta}_{rm,it}^{FF} rm_t - \hat{\beta}_{smb,it}^{FF} smb_t - \hat{\beta}_{hml,it}^{FF} hml_t \quad (6)$$

$$\hat{\alpha}_{it}^C \equiv r_{it} - \hat{\beta}_{rm,it}^C rm_t - \hat{\beta}_{smb,it}^C smb_t - \hat{\beta}_{hml,it}^C hml_t - \hat{\beta}_{pr1y,it}^C pr1y_t \quad (7)$$

This two-stage procedure yields a total of 207,968 monthly risk-adjusted before-expense returns corresponding to 3,146 different funds through 444 months. While the annualized average monthly return before expenses equals 10.17%, subtracting the risk-free rate and the part of fund returns explained by the portfolio's exposure to Fama-French three factors yields an annualized average monthly alpha of 9 basis points (bp), which is further reduced to -87 bp when momentum is taken into account. The corresponding annualized standard deviations are 19.72%, 8%, and 7.92%, respectively.

IV. A Test of the Equilibrium Predictions

In this section, we investigate whether the relation between fund performance and fees is as predicted by the market equilibrium equation (4). To do so, we first estimate the relation between mutual fund performance and expense ratios for our whole sample and then conduct a number of tests to check the robustness of the results.

A. The Relation between Performance and Fund Expenses

We first estimate equation (5) using the expense ratio as our measure of mutual fund fees. A test of the equilibrium relation between mutual fund fees and performance can be carried out by regressing performance on expenses—as specified in equation (5)—or, alternatively, by running

a regression of expense ratios on fund performance. We opt for the former approach for two reasons. The first reason has to do with comparability of results, since a number of studies have regressed different measures of performance—typically net of expenses—on expense ratios (e.g., Carhart, 1997; Chevalier and Ellison, 1999). The second reason has to do with the statistical properties of the coefficient estimates. Since funds’ true alphas are not observed, estimated alphas are used instead, so our measure of fund performance contains a significant amount of measurement error. Therefore, if performance is included as a regressor, its estimated coefficient will be biased towards zero because of the attenuation bias induced by measurement error. Regressing estimated performance on expenses (for which we expect measurement error to be much smaller), however, yields an unbiased estimate (Levi, 1973).

Our regression equation is, therefore:

$$\hat{\alpha}_{it} = \delta_{0t} + \delta_1 f_{it} + \xi_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T_i \quad (8)$$

where f_{it} stands for the fund’s expense ratio, and the intercept is allowed to vary over time to adjust for cross-sectional correlation of residuals. We estimate the model coefficients by pooled OLS and compute White’s heteroscedasticity-robust standard errors clustered by time.¹³ It is worth noting that reported standard errors are higher than OLS standard errors and, more generally, higher than standard errors computed without accounting for cross-sectional correlation of residuals across funds. Therefore, our tests for the significance of coefficient estimates will tend to be conservative.

Table II reports regression results estimated using the whole sample of diversified equity funds. When alpha is measured according to the Fama-French three-factor model, the estimated regression coefficient is both negative and statistically significant at the 10% level. In particular, investing in a fund with a one percent higher annual expense ratio reduces expected annual alpha before expenses by 62 bp. The negative relation between fees and before-fee performance is even more severe when the momentum factor is taken into account, that is, when performance is measured according to the four-factor model. In the latter case, funds with expense ratios one percent above average can be expected to earn a risk-adjusted return before expenses one percent below that of the mean fund. This effect is significant at the 1% level.

We have checked whether the results in Table II are robust both to the number of monthly

periods used for estimating fund betas and to the number of periods over which subsequent performance is measured. More specifically, we have run time series regressions using returns from the previous three years (at least 30 months of data were required). We have also aggregated alphas and expenses over the subsequent 3, 6, 12, and 24 months. In all cases, we obtain similar results both in terms of estimated coefficients and standard errors.¹⁴

The results in Table II are inconsistent with the predictions of the model sketched in Section I. Therefore, we can reject the hypothesis that the market for equity mutual funds resembles a frictionless competitive market in which net performance is equalized across funds in equilibrium.

To understand the significance of the results on Table II, it is important to note that, once we depart from the competitive benchmark, it is not clear whether we should a priori expect δ_1 to be greater or smaller than one. Thus, it seems plausible that, in the presence of costly search or other market imperfections, better funds may charge higher fees ($\delta_1 > 0$), but that those fees may not be high enough to fully compensate for differences in before-fee performance. In this scenario, funds with higher fees would offer a higher after-fee performance and the estimated δ_1 would be greater than one ($\delta_1 > 1$ implies that increases in fees are matched by larger increases in before-fee performance). It is, however, also plausible that better funds may overcharge for their ability to generate returns, leading to differences in fees that exceed differences in performance and to an estimated $\delta_1 \in (0, 1)$. In this scenario, funds with higher fees would exhibit better before-fee performance but worse after-fee performance. An extreme version of this hypothesis is the possibility that fees are completely unrelated to funds' before-fee performance, leading to $\delta_1 = 0$.

The estimated negative δ_1 in Table II, however, suggests an a priori much less plausible scenario in which before-fee performance is decreasing in fees, or, in other words, a scenario in which funds with worse before-fee performance charge higher fees. Surprisingly, this unexpected relation has been largely overlooked by the vast literature on mutual fund performance, as have been its implications regarding the functioning of the mutual fund market.

B. Is There Really a Negative Relation between Before-Fee Performance and Fees?

In this section, we evaluate the robustness of the estimated negative relation between performance and fees. In particular, we examine the possibility that it may be due to different sample

selection problems or to an incorrect measurement of the actual price paid by investors.

B.1. Index and Institutional Funds

Gruber (1996) shows that, on average, actively-managed funds underperform passively-managed index funds. Since index funds tend to charge lower fees, the apparent puzzle of a negative relation between fees and performance might be explained away by the better-known puzzle identified by Gruber (1996). In other words, fees could adjust to equalize after-fee returns among actively-managed funds, yet a negative relation between fees and before-fee performance could emerge because of investors' preference for actively-managed funds, which are more expensive and have a worse performance than index funds. In fact, in our sample, the annualized average monthly Carhart's alpha for index retail funds was -27 bp with an average expense ratio of 83 bp. This is in contrast to actively managed retail funds that delivered, on average, an annualized monthly before-expense alpha of -85 bp and had expenses of 1.46 percent. A similar argument could apply to funds sold to institutional investors. If these investors are more knowledgeable and have a greater bargaining power, it is conceivable that institutional funds may yield better performance and at a lower price. To assess these explanations, we reestimate equation (8) for the sample of retail actively-managed funds that remains after we remove both institutional and passively managed from the initial sample. As Table III shows, the estimated relation between fees and performance is still negative and highly significant for this sample. In fact, the estimated regression coefficients are marginally higher for both the three-factor and four-factor models. Therefore, the negative relation between fees and performance is not due to the influence of index or institutional funds. Throughout the rest of the paper, we nonetheless focus on actively managed retail funds to avoid any potential confounding effects of the presence of index and institutional funds.

B.2. Small and Young Funds

Another source of concern is the possibility that our results are due to the influence of funds with negligible market share, which may exhibit both low performance and high fees. Our requirement that funds have at least 48 months of return information to be included in the sample, however, already filters out the effect of unsuccessful funds that are terminated before reaching that threshold.

To ensure that our results are not driven by small funds which have survived for at least 5 years, we reestimate equation (8) for three different samples that exclude those fund-month observations with assets under management in the first decile, first and second deciles, and three first deciles of each corresponding month, respectively. Table IV shows that the negative relation between expense ratios and performance holds also for the funds with a large amount of assets under management, although the relation is not statistically significant for the measure of performance based on the three-factor model. The difference between the results obtained for the three- and four-factor models suggests that more expensive funds exhibit, on average, a greater exposure to the momentum factor.

It is worth noting that the requirement that funds have at least 48 months of return information to be included in the sample, while eliminating potential distortions due to short-lived, unsuccessful funds, also limits the representativeness of the results to the subset of seasoned funds. Although we have also estimated risk-adjusted returns using only 30 months of return information and have obtained similar results, our empirical strategy does not allow us to generalize our conclusions to the whole population of equity mutual funds. We take this as a limitation of our results, but do not attempt to extend the analysis to the youngest funds because of known problems with the data for these funds. First, Elton et al. (2001) have cautioned about the accuracy of return information for small funds. More importantly, Elton et al. (2001) have also documented a selection bias in the CRSP database, which they label omission bias: these authors highlight that a significant fraction of small funds do not have monthly information on returns and they estimate that funds with reported returns outperform those that do not report returns. Since a large fraction of young funds in our sample are also small, and since the omission bias may be especially acute for young funds,¹⁵ including young funds with return information in the analysis would introduce a selection bias, which could be problematic since young funds tend to charge higher expenses. Inspection of our dataset indeed shows that selection may be an issue for young funds, since a large fraction of observations for these funds do not include information on returns or expenses, with the incidence of omitted information being specially large for the earlier years of the sample.¹⁶

B.3. Other Fund Fees

So far, we have considered expense ratios as the only explicit cost of delegated portfolio management. As discussed in Section II, however, investors often pay other fees at the time of purchasing and/or redeeming their mutual fund shares. Hence, the previous regressions could be capturing a negative relation between performance and a specific component of total fund share ownership cost, but not necessarily a negative relation between performance and the total fees paid by investors. In particular, if more expensive funds (when only expenses are considered) charged lower loads, then after-fee performance (when all fees are considered) could still be equalized across funds. One way to circumvent this problem is to focus on no-load funds exclusively, for which annual operating expenses account for 100% of all fees. In Table V, we run the regressions for no-load funds only, which implies that about two thirds of all observations are lost. For the three-factor model, although the estimated coefficient is similar to that found above, the relation is no longer statistically significant. For the four-factor measure of performance, however, we can safely assert that performance and fees are negatively related in the no-load segment of the market.

Since load funds constitute the majority of the market, we also attempt to estimate the relation between performance and a measure of total fund share ownership cost for these funds. Following Sirri and Tufano (1998) and a large number of other studies, we compute the total annual ownership cost by adding annual expense ratios to total loads divided by the number of years investors keep their money in the fund, which we denote by τ . Although it has become common in the literature to set $\tau = 7$, redemption rates for equity funds in the period 1985-2003, suggest a shorter average holding period in the range of 2.5 to 5 years.¹⁷ We remain agnostic about τ and perform the analysis for $\tau = 2, 7$, and 10 years. As seen in Table VI, higher total ownership cost is negatively and significantly associated with before-fee performance for all the holding periods considered.

B.4. Analysis by Subperiod and by Investment Objective

To assess the temporal stability of the estimated relation between fees and performance, we divided the sample into four periods and estimated equation (5) separately for each one. As Table VII shows, the relation between the four-factor alpha and expenses is negative in all subperiods, although not significantly different from zero in the 1967-1976 period. When performance is esti-

mated using the three-factor model, the coefficient for expenses is negative and significant in the last two periods, but it is not significantly different from zero in the periods 1967-1976 and 1977-1985. Differences in results between the two measures of performance again suggest that more expensive funds have on average a greater exposure to the momentum factor.

As a final test of the robustness of our results, we divide the sample into subsamples according to funds' stated investment objective and run separate regressions for each subsample. The rationale for this strategy is that, if the mutual fund market is segmented, then certain investment objectives could exhibit both higher average fees and lower average performance while still attracting the money of investors who opt for that investment strategy. This could lead to a negative relation between fees and performance and, at the same time, be compatible with equilibrium if investors differed in their preferences over the different investment objectives (for example if they were limited in their ability to diversify). Market segmentation could also arise if investors made investment decisions sequentially by choosing first an investment objective and then a specific mutual fund within that objective.

For the 1992-2003 period (for which the classification is detailed and consistent), we divide the sample into subsamples according to the fund's *Strategic Insight* objective code as reported by CRSP, and then run the regression for each subsample. As shown in Table VIII, expense ratios are negatively related to performance in all five investment objectives, although the relation is not statistically significant for *Growth MidCap* and *Small Company Growth* funds. Replacing expense ratios with total ownership cost with $\tau = 2, 7$, and 10 years, produces results similar to those of Tables VII and VIII (not reported).

V. Explaining the Puzzle

A. Cost-Based Explanations

We consider two different kinds of explanations for the negative relation between before-fee performance and fees, which differ in their assumptions regarding mutual funds' pricing behavior. According to the first type of explanation, fees reflect funds' costs of operating the fund. If low costs were associated with better before-fee performance, a univariate regression would result in a negative relation between fees and performance.

Although one may expect performance to be positively, rather than negatively, associated with fund costs—if higher costs reflect larger investments in research tools or higher salaries to attract more talented managers—, there are also plausible explanations for a negative correlation between costs and performance. For instance, it is likely that there exist economies of scale in fund management. At the same time, a larger size could be associated with better performance if a fund’s size reflects its past performance, and performance is persistent. Similarly, older funds may benefit from learning economies, which may be passed on to investors in the form of lower fees. If fund longevity were related to good performance—as would be the case if worse-performing funds were more likely to close down—then, again, we could observe a negative relation between costs and performance.¹⁸ Finally, some funds could just be run more efficiently than others, with the differences in the quality of fund management manifesting themselves both in terms of higher returns and lower operating costs.

Therefore, the negative relation between before-fee risk-adjusted returns and fees that resulted from the univariate analysis conducted in Section IV could simply be due to a failure to control for funds’ operating costs.

B. Strategic-Pricing Explanations

The second type of explanation views the negative relation between before-fee performance and fees as the result of strategic fee setting by mutual fund companies. The challenge for this kind of hypothesis is to explain why fund managers with worse past or expected performance may strategically decide to set high fees. A possible explanation has been proposed and empirically tested for money market mutual funds by Christoffersen and Musto (2002). Christoffersen and Musto (CM, hereafter) argue, on the basis of empirical studies on mutual fund investment flows (e.g. Sirri and Tufano, 1998) and survey data on mutual fund investors’ behavior (Capon et al., 1996; Alexander *et al.*, 1997), that mutual fund investors differ in their performance sensitivity, with some investors quickly moving their money in response to differences in performance and others reacting much more sluggishly to those differences. Since the demand function (that relates the demand for a fund’s shares to that fund’s fee and past performance) faced by a fund is likely to be determined to a large extent by its current investors, the elasticity of that demand to performance

is likely to be largely determined by the performance-sensitivity of the current investors of the fund. Therefore, funds with a larger proportion of performance-insensitive investors will charge higher fees, since for these funds the reduction in after-fee performance caused by an increase in fees will not translate into a large flow of money out of the fund. Finally, CM argue that funds with a worse performance track record will have a less performance-sensitive clientele, since the performance-sensitive investors will have fled those funds following bad performance. It follows that funds with bad past performance will find it optimal to charge higher fees. CM's explanation is testable: all that is needed is a measure of the performance-sensitivity of each fund's flows, an issue that we address in the next subsection.

A different strategic explanation for the negative relation between before-fee performance and fees has been recently provided by Gil-Bazo and Ruiz-Verdú (2005)—GR, hereafter. Like CM, these authors argue that there are performance-sensitive and performance-insensitive investors. Then, they show that competition between funds for the money of performance-sensitive investors will lead good-quality funds to lower their fees up to a point where they effectively price bad-quality funds out of the performance-sensitive segment of the market. Unable to compete in that segment, bad funds will raise their fees to extract rents from performance-insensitive investors. The reason why good funds are able to price worse funds out of the market has to do with the way fund managers are compensated. The revenues of mutual fund managers are typically determined as a fraction of assets under management. Therefore, for any given fee (expressed as a fraction of asset value), good fund managers, who will achieve a larger increase in the value of their assets, will earn higher revenues. As a result, there exists a fee level at which good funds break even in expectation while worse-performing funds incur an expected loss.

GR's explanation for the negative relation between before-fee performance and fees differs from the one provided by CM in that, rather than responding to past returns, funds' fee strategies are forward-looking: fund managers who expect to perform poorly set higher fees and, thus, end up with the less performance-sensitive investors. This implies, in particular, that if we consider two funds with similar clienteles (in terms of the performance sensitivity of investors), the one with lower expected returns will set a higher fee. This prediction is also testable. What is needed in this case is a measure of a fund's risk-adjusted expected performance. Estimated alpha ($\hat{\alpha}_{it}$) will be a good measure of expected performance (α_{it}), as long as the measurement error in $\hat{\alpha}_{it}$ is

not correlated with the level of fees. If no such correlation exists, the result of including realized performance, rather than expected performance, as a regressor in the fee equation reduces to the well-known attenuation bias in the presence of measurement error. Thus, when interpreting the performance coefficient estimates, one should bear in mind that they will be biased toward zero.

C. Empirical Strategy

To test the empirical validity of the different explanations for the negative relation between before-fee performance and fees, we investigate how fees vary with fund characteristics. To do so, we assume that fund i 's fee at time t , f_{it} , is a function F of: a) a vector $\mathbf{x}_i = (x_{i1}, \dots, x_{iK})$ of variables that are likely to determine the fund's operating costs; b) the performance-sensitivity of the fund's flows, S_i ; and c) the fund's expected performance in period t , α_{it} .¹⁹

$$f_{it} = F(\mathbf{x}_{it-1}, S_{it}, \alpha_{it}) + \nu_{it}, \quad (9)$$

where ν_{it} is a generic error term. Since we are interested both in the total price paid by investors and in the compensation of managerial skill, we perform regressions both for total cost of ownership and for a measure of management fees defined below.

We build on the literature on the determinants of mutual fund fees,²⁰ which has mostly considered fund fees as a reflection of operating costs, to select the variables that may influence the costs of operating a fund. These variables are the following:

1. *Size.* Mutual fund management is likely to exhibit scale economies, since a significant fraction of the costs of managing a fund are fixed. We include size squared in the regression to allow for the possibility that funds may experience diseconomies of scale beyond a certain size. As discussed above, fund size may also be correlated with fund quality.
2. *Age.* Costs may fall with age if there are learning economies in fund management. Age can also be correlated with a fund's quality because of learning economies or, simply, because better funds are more likely to survive.
3. *Complex size.* Since there may be economies of scale at the management company level (Malhotra and Mcleod, 1997; Tufano and Sevick, 1997; Latzko, 1999), we also include complex

size in the regression. As with fund size, we also include the square of complex size to account for the possibility of eventual diseconomies of scale.

4. *Number of funds in complex.* A higher number of funds to manage, holding total assets constant, may increase total management costs due to the increase in resources needed to manage additional funds.
5. *Turnover.* A high turnover may signal a management strategy that requires frequent trading, with frequent trading, in turn, requiring greater management effort. Consistently with this explanation, Chalmers et al. (2000) have found transaction costs to be positively correlated with expense ratios. It is important to note that turnover has also a direct negative impact on performance through increased transaction costs (which are directly deducted from asset value). Controlling for turnover, thus, enables us to evaluate an alternative explanation to the puzzle: funds that trade too much underperform and are more costly to manage.
6. *Volatility.* The volatility of a fund's returns has also been proposed as a determinant of fund management costs, with the presumption that greater volatility signals a greater difficulty in managing the fund. As in the case of turnover, differences in volatility may also explain differences in performance across funds. It has been documented that mutual funds lagging behind their rivals tend to adopt riskier strategies (Chevalier and Ellison, 1997). Therefore, the set of high-risk funds may contain a relatively large fraction of underperformers.
7. *Investment objective.* Different investment objectives may require different amounts of research and oversight. For example, it is often argued that funds that pursue an aggressive growth investment objective will be more costly to manage than those with a growth and income objective. It is also plausible that investment opportunities may be associated with different asset classes. We, thus, include investment objective dummies to account for potentially different cost structures and different mean performance. As in Section IV, we classify funds into investment objectives using the *Strategic Insight* objective code as reported by CRSP.
8. *Fee structure.* Investors acquiring funds through brokers or financial advisers also have to bear the cost of compensating those intermediaries. Therefore, we also include a dummy variable

to identify single-class load funds and dummies for the main share classes. We include these dummies to correct for the potential distortions induced by employing a homogeneous holding period for all funds when investors with different holding periods can be expected to select different share classes.

D. The Performance Sensitivity of Fund Flows

The second main component of our fee equation is the performance sensitivity of a fund's flows. CM suggest that the performance sensitivity of a fund's investors depends on the outflows of money experienced by the fund in the past (attrition), with funds that have experienced the largest outflows being left with the least performance-sensitive investors.²¹ Consistently with this reasoning, CM propose the following measure of the elasticity of flows to performance:

$$Q/MAX_{it} = \frac{TNA_{it}}{MAX_{it}}, \quad (10)$$

where TNA_{it} is fund i 's total net asset value at the beginning of period t and MAX_{it} is the maximum total net asset value of fund i in the time-span up to period t . Q/MAX measures asset retention, so that $1 - Q/MAX$ measures asset attrition.

Although Q/MAX is a sensible measure of flow-to-performance sensitivity, there are several reasons why we need to control for other factors. First, this measure does not take into account the direct effect of returns on changes in asset value. A low value of Q/MAX may not be due to past outflows of money, but, rather, to recent low returns. Second, when $Q < MAX$, there may be factors that both increase Q/MAX and reduce performance sensitivity. In particular, Huang et al. (2006) have shown that variables that reduce investor participation costs (such as fund affiliation with large families or the presence of a star in the fund's family) are associated with larger net flows of money. At the same time, the sensitivity of flows to performance in the high-performance range is lower for funds with low participation costs. Thus, if a fund's participation costs are reduced because of, say, the appearance of a star fund in its family, Q/MAX may go up because of increased net flows, while, at the same time, the performance sensitivity of the fund's investors falls. Finally, Q/MAX is sensitive to a fund's age. Thus, funds alive during a period of high asset appreciation will be more likely to have a larger value of MAX_{it} (and a lower value of Q/MAX) in subsequent

periods than funds that were not alive during such a period.

Instead of taking Q/MAX_{it} as a proxy for flow-to-performance sensitivity, we propose to obtain a direct estimate of flow-to-performance sensitivity as a function of fund characteristics. We proceed in two steps. In the first step, we estimate the coefficients of a model of money flows into mutual funds based on extant studies of fund flows. In the second step, we obtain our measure of performance sensitivity from the estimated coefficients and fund characteristics.

D.1. Determinants of Fund Flows

As customary, we define annual net flow to fund i in month t , $Flow_{it}$, as the relative growth of the fund's total net assets (TNA) adjusted for returns net of expenses, R_{it}^n :²²

$$Flow_{it} = \frac{TNA_{it} - TNA_{it-1}(1 + R_{it}^n)}{TNA_{it-1}} \quad (11)$$

To estimate performance sensitivity, we propose a model for fund flow determination that encompasses the main stylized facts that have emerged from the literature on fund flows:

1. Investors chase past performance, i.e., flows of money to mutual funds are positively related to recent relative after-expense performance (see, among others, Gruber, 1996, and Sirri and Tufano, 1998).
2. Flows also depend on other variables, such as fund size (Sirri and Tufano, 1998), fund age (Chevalier and Ellison, 1997), fund expenses (Sirri and Tufano, 1998; Barber et al., 2005), total complex size (Nanda et al., 2005), lagged flows (Jain and Wu, 2001), and flows of money to funds with the same investment objective (Sirri and Tufano, 1998).
3. Flow-to-performance sensitivity depends on fund age with flows being less sensitive to performance for older funds (Chevalier and Ellison, 1997; Huang et al., 2006). We also test CM's hypothesis that Q/MAX increases the sensitivity of flows to performance.
4. The sensitivity of flows to performance is significantly higher for recent top performers, i.e., the flow-to-performance function is convex (Ippolito, 1992; Gruber, 1996; Chevalier and Ellison, 1997; Sirri and Tufano, 1998).

5. For funds with low participation costs, flow-to-performance sensitivity is higher in the low and medium performance range, and lower in the high performance range, i.e., the flow-performance curve becomes less convex as participation costs decrease (Huang et al., 2006).

Consistently with these findings, we estimate the following regression equation:

$$\begin{aligned}
Flow_{it} = & a_{0t} + \sum_{j=1}^J a_j x_{j,it-1} + b_0 Perf_{it-1} + b_1 Perf_{it-1} rel_age_{it-1} + \\
& + b_2 Perf_{it-1} rel_Q/MAX_{it-1} + a_H I_{H,it-1} + b_H Perf_{it-1} I_{H,it-1} + \\
& + a_H^{PC} I_{H,it-1} PC_{it-1} + b_H^{PC} Perf_{it-1} I_{H,it-1} PC_{it-1} + \varphi_{it},
\end{aligned} \tag{12}$$

where φ_{it} stands for a generic error term. The proxy for past performance, $Perf_{it-1}$, is the fund's four-factor alpha in year $t-1$, net of expenses, and in excess of the mean performance of all funds with the same investment objective in that year. $I_{H,it}$ is a dummy variable that equals one if $Perf_{it}$ is among the top third of all funds with the same investment objective in year t . We include this variable to allow for a convex relation between performance and flows. Variables rel_age_{it} and rel_Q/MAX_{it} are, respectively, the log of the fund's age in years and the fund's Q/MAX in excess of the average of those variables for all observations in year t with the same investment objective as fund i . Control variables, x_j , $j = 1, \dots, J$, include: fund size and age; front-end load; 12b-1 fee; non-marketing expenses; dummy variables for share classes; return volatility; total net asset value for all funds under the same management company; lagged flows; total flows of money to all funds with the same investment objective; and Q/MAX .

Variable PC is a proxy for participation costs. We consider two of the proxies proposed by Huang et al. (2006): total assets managed by the company (in excess of the category's average on that year); and a dummy for the presence of a "star" fund managed by the management company. More precisely, this dummy variable equals one for fund i if there is another fund managed by fund i 's management company with performance in the top 5% of its category.²³

Table IX shows estimated coefficients of equation (12). The table is largely consistent with previous results. The positive and significant values of a_H and b_H show that top-performing funds receive higher flows on average, and that flows into those funds are more sensitive to relative performance, consistently with the extensively documented convexity of the flow-performance relation. Lack of sensitivity of flows to past performance in the low performance range is also consistent with

the existence of unsophisticated or disadvantaged investors (although see Lynch and Musto, 2003, for an alternative interpretation of this lack of sensitivity consistent with investor rationality).

Larger funds grow relatively less, while being part of a large complex or belonging to a fast-growing investment objective category leads to larger inflows of money. Regarding age, we find that older funds tend to experience lower net flows and, as in previous studies, our estimation results indicate that investors are less responsive to fund performance if the fund is older. Also in line with prior studies, fund flows appear to be persistent. Consistently with Huang et al. (2006) the flow-to-performance sensitivity among top performing funds decreases when fund investors face lower participation costs, regardless of the proxy for participation costs employed. Finally, Table IX also shows that high values of Q/MAX are significantly associated with higher flow-to-performance sensitivity, supporting CM's hypothesis that Q/MAX is associated with performance sensitivity.

D.2. Performance Sensitivity

Once we have estimated the coefficients in (12), we obtain our proxy for flow-to-performance sensitivity as the first derivative of flow with respect to performance:

$$\begin{aligned} S_{it} &= \frac{\partial E_{t-1}(Flow_{it})}{\partial Perf_{it-1}} = \\ &= \hat{b}_0 + \hat{b}_1 rel_age_{it} + \hat{b}_2 rel_Q/MAX_{it} + \hat{b}_H I_{H,it-1} + \hat{b}_H^{PC} I_{H,it-1} PC_{it-1} \end{aligned} \quad (13)$$

where $E_{t-1}(\cdot)$ denotes the expectation operator conditional on the information set at time $t - 1$, and the hats denote estimated coefficients.

E. Performance, Operating Costs and Fund Fees

With the measure of performance sensitivity derived above, we estimate the following linear specification of fee equation (9):

$$\begin{aligned}
f_{it} = & \gamma_{0t} + \gamma_1 size_{it-1} + \gamma_2 size_{it-1}^2 + \gamma_3 age_{it-1} + \gamma_4 size_complex_{it-1} + \gamma_5 size_complex_{it-1}^2 + \\
& + \gamma_6 n_complex_{it-1} + \gamma_7 turn_{it-1} + \gamma_8 \sigma_{it-1} + \gamma_9 classA_{it-1} + \gamma_{10} classB_{it-1} + \\
& + \gamma_{11} classC_{it-1} + \sum_{j=1}^4 \lambda_j obj_{it-1}^j + \gamma_{12} S_{it} + \gamma_{14} \hat{\alpha}_{it} + \nu_{it}
\end{aligned} \tag{14}$$

Since fee data are available yearly during most of the sample period, the time index t in equation refers to calendar years (fee information is reported quarterly only for the last three years of the sample). In equation (14), $size_{it}$ is defined as the log of fund i 's total net asset value averaged over year t (starting in year 2000, total net assets are reported quarterly); age is the log of the fund's age in years; $size_complex_{it}$ and $n_complex_{it}$ are defined as the sum of $size$ over all funds managed by the management company that manages fund i , and the total number of funds managed by that company, respectively; $turn_{it}$ is the reported portfolio turnover averaged over year t ; and obj_{it}^j is a dummy variable that takes value one if the fund's investment objective is j . The performance measure, $\hat{\alpha}_{it}$, is the fund's cumulative risk-adjusted return during year t , and is computed as the sum of monthly alphas. Finally, σ_{it} is the standard deviation of the fund's monthly returns in year t . As in the estimation of fund flows, we restrict our analysis to the period 1992-2003.

Table X reports estimation results for different specifications of the fee equation (14). In columns (1)-(2), the dependent variable is total ownership cost, computed assuming a holding period of 7 years, while in column (3), the dependent variable is management fees, defined as expenses net of 12b-1 fees.²⁴ For funds with a single share class, total ownership cost is computed by adding annual expense ratios to total loads amortized with $\tau = 7$. We also assume that for this holding period, all class B and C shares are exempt from contingent deferred sales charges. Therefore, for those shares we compute the total ownership cost as the sum of the expense ratio and the front-end load amortized with $\tau = 7$. Although it might be argued that investors with a holding period of 7 years will not consider investing in class A shares, we consider all share classes in the regression, but control for differences in costs by including share class dummies. We also include a dummy

variable that equals one for load funds with a single share class. As in Section IV, we report pooled OLS regression results with time dummies and standard errors clustered by time (year).

Results regarding those variables identified as determinants of funds' operating costs are in line with previous studies of mutual fund fees. Across all specifications of the fee equation, results confirm the existence of economies of scale—both at the fund and management company levels—in the management of mutual funds that are passed on to investors in the form of lower fees. The negative coefficient for the square of size is also evidence that there may be diseconomies of scale for large enough funds. Also consistently with previous evidence, both high portfolio turnover and volatility are associated with higher total ownership cost. Finally, results also show significant differences in fees across investment categories, with *Aggressive Growth* funds (the excluded dummy) being as much as 15 bp per year more expensive than the cheapest category: *Growth and Income*. Taken together, these estimated coefficients suggest that operating costs play an important role in determining fund fees.

Column (1) in Table X presents the results of estimating equation (14) without the performance-sensitivity measure. If the negative relation between expected performance and fees were the consequence of the omission of variables—such as size, age or turnover—that are likely both to determine operating costs and to be related to performance, then we would expect the coefficient of expected performance to change sign, or, at least, become statistically insignificant once these variables are included in the regression. As column (1) shows, this is not the case: the coefficient for expected performance remains negative and significantly different from zero. Cost-based arguments, therefore, cannot explain why fees and performance are negatively related.

To evaluate the extent to which the effect of performance on fees operates through differences in investors' performance-sensitivity correlated with differences in performance, column (2) reports the results of estimating the full model (14), which includes both performance and performance-sensitivity as regressors.²⁵ The results in column (2) show that performance sensitivity does have a negative, and significant at any reasonable significance level, effect on fees, a result that suggests that equity mutual funds strategically exploit a low elasticity of demand with respect to net performance to increase their fees. Therefore, our results extend the findings of CM—which were obtained for a cross-section of money market mutual funds—to the market for actively-managed equity mutual funds, for a much larger sample, and with a more precise measure of performance

sensitivity.²⁶ The inclusion of a measure of performance sensitivity, however, does not eliminate the negative effect of expected performance on fees. On the contrary, the estimated coefficient in column (2) is not only negative and statistically significant but higher (in absolute value) than the one in column (1). In other words, elasticity of demand appears to be an important determinant of fees, but does not by itself explain why underperforming funds set higher fees. Therefore, the results lend support to the hypothesis that funds that expect to have a lower performance charge higher fees.

The above results show that the shares of underperforming funds are more expensive. From this finding, however, one cannot conclude that the managers of underperforming funds are being paid more per dollar under management than the managers of better-performing funds. The reason is that their higher cost could be due to the effect of distribution costs. To investigate the pricing of fund management skills, we estimate equation (14) with management fees (measured as expenses minus 12b-1 fees) as the fee variable. The results of this regression, reported in column (3) are very similar to the ones obtained for total ownership cost. Therefore, our estimates indicate that underperforming managers are charging a higher price for their fund management services than better-performing ones.

A potential problem with our interpretation of the results in Table X is one of reverse causality. Namely, the estimated coefficient may not reflect the effect of expected performance on current fees, but, rather, the effect of current fees on expected performance. However, to the extent that current fees have an effect on next year's performance, we would expect this effect to be positive, since higher fees increase the resources available to fund managers and, thus, can be expected to generate higher returns. Therefore, if this sort of reverse causality is operating, we would expect it to work against finding a negative coefficient for performance.

There is another reason why the results of Table X should be interpreted with care. Our finding that funds setting higher fees have worse before-fee performance does not necessarily imply that investors should avoid expensive funds, since the higher fees charged by those funds could be used to offer services (such as investment advice, the provision of frequent and clear statements, or the availability of agents that can respond to investors' inquiries) that are valued by investors. Therefore, our results could be due to a differentiation strategy by mutual fund companies, which could target their better-performing funds to sophisticated investors and offer a bundle of lower

performance and better service to investors who value that service more. Our data does not allow us to reject this hypothesis, but there are indications that it may, at best, explain only part of the differences in fees. First, the estimated coefficient for the fee-performance relation does not vary substantially when we control for size and size of the management company, although we would expect larger management companies to offer a higher level of service. And second, the size of the coefficient is not reduced when we replace total ownership cost (which include fees that are devoted to paying for the investment advice provided by brokers and other distribution costs) by our management fee proxy in the regression.

VI. Conclusions

In this paper, we show that there exists a negative relation between funds' before-fee performance and the fees they charge to investors. Since this evidence is at odds with economic intuition—and the predictions of a benchmark competitive model—we subject it to a battery of robustness tests, and find that it survives all of them.

We then propose two kinds of explanations for this anomalous result. According to the first kind of explanation, the negative relation is spurious and results from the fact that there are factors, omitted in univariate regressions, that are both positively correlated with returns and negatively correlated with funds' operating costs and, thus, also with funds' fees. According to the second kind of explanation, in contrast, the negative relation is the result of funds' strategically setting fees as a function of their past or expected performance. In particular, we consider two rationales for this kind of strategic behavior. The first one, proposed by Christoffersen and Musto (2002), argues that funds with worse past performance have a pool of investors that are not very sensitive to fund performance. Faced with an inelastic demand for their shares, underperforming funds optimally increase fees. The second explanation, proposed by Gil-Bazo and Ruiz-Verdú (2005), argues that funds with low expected performance optimally set high fees and target performance-insensitive investors, since they anticipate that they will not be able to compete with better-performing funds in the market for sophisticated investors. Better-performing funds, on the other hand, keep fees low because of competition among them for the money of performance-sensitive investors. The empirical analysis finds support for the two strategic-pricing explanations. Even though funds'

operating costs are important determinants of fees, they do not explain away the negative relation between before-fee performance and fees. Our regression analysis shows that, controlling for those cost determinants, funds with lower expected performance and funds with less elastic demands charge higher fees.

In the process of obtaining our results, we bring together three different strands of research: the empirical investigation of fund performance, the analysis of fund flows, and the study of the determinants of fund fees. Although there exist performance studies that look at the effect of funds' expenses, and fee studies that include performance as a regressor, the main contribution of this paper is to provide a thorough examination of the puzzling relation between the two variables.

Our results have important consequences both for investors and for regulators. For the former, they provide further evidence that high fund fees do not generally pay for commensurately high risk-adjusted returns. On the contrary, high fees may be a signal of poor performance. The implications for regulators are wide-reaching. First, we provide further evidence that a significant fraction of mutual fund investors is insensitive to fund fees and after-fee performance, in line with the results of some recent studies (e.g., Gruber, 1996; Sirri and Tufano, 1998; Elton et al., 2004) and existing survey and experimental evidence on mutual fund investors' behavior (Capon et. al., 1996; Alexander et. al., 1997; Choi et. al., 2006). At least for the last decade of our sample, this lack of performance-sensitivity coexists with easily available public information about funds' fees and performance, and with an increasing recognition by the financial press of the fact that fund fees do not generally buy higher returns.²⁷ Therefore, our results argue for the design of mutual fund disclosure requirements that take into account the cognitive limitations of investors. They also point out a potential drawback of the complete delegation of retirement investment decisions to individual investors. Second, our results show that competition in the mutual fund market does not guarantee that fund management services are adequately priced. On the contrary, there exists a perverse negative relation between fund performance and the price of fund management that allows underperforming funds to survive and that, thus, may also provide wrong incentives for entry.

References

- Alexander, G. J., J. D. Jones, and P. J. Nigro, 1997, Mutual fund shareholders: Characteristics, investor knowledge, and sources of information, *Office of the Comptroller of the Currency, E&PA Working Paper* 97–13.
- Barber, B. M., O. Terrance and L. Zheng, 2005, Out of sight, out of mind: the effects of expenses on mutual fund flows, *Journal of Business* 78, 2095–2120.
- Berk, J. B. and R. C. Green, 2004, Mutual fund flows and performance in rational markets, *Journal of Political Economy* 112, 1269–1295.
- Brown, S. and W. Goetzmann, 1995, Performance persistence, *Journal of Finance* 50, 679–698.
- Brush M., 1997, How fund costs drag down gains, *Money*, July 1.
- Busse, J. and P. Irvine, 2006, Bayesian alphas and mutual fund persistence, *Journal of Finance* 61, 2251–2288.
- Capon, N., G. Fitzimmons, and R. Prince, 1996, An individual level analysis of the mutual fund investment decision, *Journal of Financial Services Research* 10, 59–82.
- Carhart, M.M., 1997, On persistence in mutual fund performance, *Journal of Finance* 52, 57–82.
- Carhart, M.M., J.N. Carpenter, A.W. Lynch, and D.K. Musto, 2002, Mutual fund survivorship, *Review of Financial Studies* 15, 1439–1463.
- Chalmers, J.M.R., R.M. Edelen, G.B. Kadlec, 2001, On the perils of security pricing by financial intermediaries: The wildcard option in transacting mutual fund shares, *Journal of Finance* 56, 2209–2236.
- Chen, J., H. Hong, M. Huang, and J.D. Kubik, 2004, Does fund size erode mutual fund performance? The role of liquidity and organization, *The American Economic Review* 94, 1276–1302.
- Chevalier, J. and G. Ellison, 1997, Risk taking by mutual funds as a response to incentives. *Journal of Political Economy* 105, 1167–1200.
- Chevalier, J. and G. Ellison, 1999, Are some mutual fund managers better than others? Cross-sectional patterns in behavior and performance, *Journal of Finance* 54, 875–899.

- Choi, J. J., D. Laibson and B. C. Madrian, 2006, Why does the law of one price fail? An experiment on index mutual funds, *NBER Working Paper* 12261.
- Christoffersen, S. and D. Musto, 2002, Demand curves and the pricing of money management, *Review of Financial Studies* 15, 1499–1524.
- Cohen, R., J. Coval, and L. Pastor, 2005, Judging fund managers by the company they keep, *Journal of Finance* 60, 1057–1096.
- Daniel, K., M. Grinblatt, S. Titman, and R. Wermers, 1997, Measuring mutual fund performance with characteristic-based benchmarks, *Journal of Finance* 52, 1035–1058.
- Deli, D. N., 2002, Mutual fund advisory contracts: An empirical investigation, *Journal of Finance* 57, 109–133.
- Elton, E., M. Gruber, and C. Blake, 1996, The persistence of risk-adjusted mutual fund performance, *Journal of Business* 69, 133–157.
- Elton, E., M. Gruber, and C. Blake, 2001, A first look at the accuracy of CRSP mutual fund database and a comparison of the CRSP and Morningstar mutual fund databases, *Journal of Finance* 56, 2415–2430.
- Elton, E., M. Gruber, and C. Blake, 2003, Incentive fees and mutual funds, *Journal of Finance* 58, 779–804.
- Elton, E., M. Gruber, and J. A. Busse, 2004, Are investors rational? Choices among index funds, *Journal of Finance* 59, 261–288.
- Evans, R. B., 2006, Does alpha really matter? Evidence from mutual fund incubation, termination and manager change, *mimeo*.
- Fama, E. and K. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.
- Fama, E., and J. MacBeth, 1973, Risk, return and equilibrium: Empirical tests, *Journal of Political Economy* 81, 607–636.
- Ferris, S., and M. Chance, 1987, The effect of 12b-1 plans on mutual fund expense ratios: A note, *Journal of Finance* 42, 1077–1082.

- Gil-Bazo, J. and P. Ruiz-Verdú, 2005, When cheaper is better: Fee determination in the market for equity mutual funds, *Department of Business Administration, Universidad Carlos III de Madrid, Working Paper 05-43, Business Economics Series 09*.
- Golec, J., 2003, Regulation and the rise in asset-based mutual fund management fees, *Journal of Financial Research* 1, 19–30.
- Gruber, M., 1996, Another puzzle: The growth in actively managed mutual funds, *Journal of Finance* 52, 783–810.
- Hadi, A.S., 1994, A modification of a method for the detection of outliers in multivariate samples, *Journal of the Royal Statistical Society, Series B-Methodological* 56, 393–396.
- Hendricks, D., J. Patel, and R. Zeckhauser, 1993, Hot hands in mutual funds: short-run persistence of relative performance, 1974-1988, *Journal of Finance* 48, 93–130.
- Huang, J., K.D. Wei, and H. Yan, 2006, Participation costs and the sensitivity of fund flows to past performance, *Journal of Finance* (forthcoming).
- Investment Company Institute, 2006, *Mutual Fund Factbook, 46th Edition* (Washington, D.C.: Investment Company Institute).
- Ippolito, R. A., 1992, Consumer reaction to measures of poor quality: Evidence from the mutual fund industry, *Journal of Law and Economics* 35, 45–70.
- Jain, P.C., and J.S. Wu, 2001, Truth in mutual fund advertising: Evidence on future performance and flows, *Journal of Finance* 55, 937–958.
- Jensen, M. C., 1968, The performance of mutual funds in the period 1945-1964, *Journal of Finance* 23, 389-416.
- Kacperczyk, M., C. Sialm, and L. Zheng, 2005, On the industry concentration of actively managed equity mutual funds, *Journal of Finance* 60, 1983–2011.
- Kacperczyk, M. and A. Seru, 2006, Fund manager use of public information: New evidence on managerial skills, *Journal of Finance* (forthcoming).
- Kosowski, R., A. Timmermann, R. Wermers and H. White, 2006, Can mutual fund “stars” really pick stocks? New evidence from a bootstrap analysis, *Journal of Finance* (forthcoming).

- Kothari, S., and J. Warner, 2001, Evaluating mutual fund performance, *Journal of Finance* 56, 1985–2010.
- Latzko, D. A., 1999, Economies of scale in mutual fund administration, *Journal of Financial Research* 22, 331–339.
- Levi, M. D., 1973, Errors in the variables bias in the presence of correctly measured variables, *Econometrica* 41, 985–986.
- Luo, G., 2002, Mutual fund fee-setting, market structure and mark-ups, *Economica* 69, 245–271.
- Lynch, A. W. and D. K. Musto, 2003, How investors interpret past fund returns, *Journal of Finance* 58, 2033–2058.
- Mahoney, P. G., 2004, Manager-investor conflicts in mutual funds, *Journal of Economic Perspectives* 18, 161–182.
- Malhotra, D. and R. McLeod, 1997, An empirical analysis of mutual fund expenses, *Journal of Financial Research* 20, 175–190.
- Malkiel, B., 1995, Returns from investing in equity mutual funds 1971 to 1991, *Journal of Finance* 50, 549–572.
- Nanda, V., Z. Wang, and L. Zheng, 2005, The ABCs of mutual funds: A natural experiment on fund flows and performance, *University of Michigan, Working paper*.
- Petersen, M., 2006, Estimating standard errors in finance panel data sets: Comparing approaches, *Northwestern University, Kellogg School of Management, Working Paper*.
- Ross, Stephen A., 1976, The arbitrage theory of capital asset pricing, *Journal of Economic Theory* 13, 341–360.
- Sharpe, W. F., 1966, Mutual Fund Performance, *Journal of Business* 39, 119–138.
- Sirri, E. and P. Tufano, 1998, Costly search and mutual fund flows, *Journal of Finance* 53, 1589–1622.
- Tufano, P. and M. Sevick, 1997, Board structure and fee-setting in the U.S. mutual fund industry, *Journal of Financial Economics* 46, 321–355.
- Wermers, R., 2000, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transaction costs, and expenses, *Journal of Finance* 55, 1655–1703.

Data Appendix

A. Investment Objectives

The CRSP database offers different classifications of funds by investment objective. Unfortunately, no classification covers the whole 1962-2003 period. To construct our sample, we therefore had to combine the information provided by the different classifications. We considered a fund to be a domestic diversified equity mutual fund if it satisfied any of the following conditions:

- For the years 1962-1989 a fund was included in the sample if the type of securities mainly held by the fund was *common stocks* and the fund belonged to any of the following Wiesenberger Objective codes: *Growth*, *Growth-Income*, *Maximum Capital Gain*.
- For the years 1990-1991, there is no information on the type of security mainly held by the fund and the Wiesenberger classification changes. For these years, a fund was included in the fund if it belonged to any of the following Wiesenberger Objective codes: *Growth and Current Income*, *Long-Term Growth*, *Maximum Capital Gains*, or *Small Capitalization Growth*.
- For the years 1992-2003, a fund was included in the sample if it belonged to any of the following *Strategic Insight* investment objective categories: *Aggressive Growth*, *Growth Mid Cap*, *Growth and Income*, *Growth*, *Small Company Growth*.

B. Index Funds

We coded as index funds those whose name contained any of the following strings: Index, Idx, Ix, Indx, NASDAQ, Nasdaq, Dow, Mkt, DJ, S&P, 500, BARRA. We checked the accuracy of our variable using information for the year 2004 from Standard & Poors Fund Services (the information was freely provided at Standard & Poors web site and retrieved July 14-15, 2004). The correlation between our index dummy and Morningstar's specialty code "index" is 0.72 and our dummy captures 70% of the observations coded as index by Morningstar (at the same time, 22% of the observations coded by our dummy as index are not classified as such by Morningstar). Given the time lag (some of the funds in our sample were not alive a year later) and the difference between the datasets, we are confident that our dummy variable is capturing most index funds, without unnecessarily excluding non-index funds.

C. Institutional Funds

We coded as institutional funds (or share classes) those funds whose name contained the strings “Inst” or “inst” and those that belonged to share classes “Y” or “I”. We compared our coding for the year 2003 with the one provided by Morningstar for the year 2004 (obtained from the MSN Money webpage on July 14, 2004). Despite the one year difference, the correlation between the two codings was 0.67; the percentage of the observations coded as institutional by Morningstar that was also coded as such by our measure was 67% (at the same time, 22% of the observations coded by our dummy as institutional were not classified as such by Morningstar).

D. Share Classes and Fund Complexes

In the CRSP dataset, different classes of the same fund appear as different funds without a common fund identifier. Since funds with different classes are named “*Fund’s name/Class*,” we performed a name search to identify fund-classes belonging to the same fund. The dataset, however, does include an identifier of the company that manages the fund. We employ this identifier to compute all variables related to fund complexes.

Notes

¹The literature on mutual fund performance is vast with first entries going back to Sharpe (1966) and Jensen (1968). More recent entries to this literature include Malkiel (1995); Brown and Goetzmann (1995); Gruber (1996); Carhart (1997); Daniel et al. (1997), Hendricks et al. (1993); Elton et al. (1996); Wermers (2000); Cohen et al. (2005); Kacperczyk et al. (2005); Busse and Irvine (2006); Kosowski et al. (2006).

²Fund expenses include the management fee paid to the fund management company. In Section II, we describe the fees charged by mutual funds in greater detail.

³In 2005, U.S. households held about 20 percent of their assets—excluding real estate and other property—in mutual funds, and nearly half of those households owned mutual funds (Investment Company Institute, 2006).

⁴Recently, Berk and Green (2004) have provided a partial equilibrium model of the mutual fund market that also requires that, in equilibrium, all funds' risk-adjusted excess returns be zero.

⁵Some funds allow the percentage to depend on the fund's performance. Our data does not allow us to identify whether a fund charges a fixed or a variable (or *incentive*) fee. However, it follows from the evidence reported in Elton et al. (2003) that for most of our sample period, the fraction of funds with incentive fees is very small. Therefore, we simply assume throughout the paper that all funds charge fixed fees.

⁶Back-end loads (*contingent deferred sales charges*) are often computed as a fraction of the minimum of the value of the shareholders initial investment and the value of the shareholders investment at the time of redemption. Often, the percentage charged to the investor depends on the time the investor has held the fund's shares, and the sales charge may be waived if the investor holds the shares for a long enough period. Mahoney (2004) provides a review of mutual fund fee practices and regulation.

⁷About 0.8% of observations were identified as outliers.

⁸See the appendix for the procedure we followed to group fund-classes into funds.

⁹Wermers (2000), Kothari and Warner (2001), Kacperczyk et al. (2005) and Kosowsky et al. (2006), are only a few recent examples of papers employing Carhart's model to measure mutual fund performance.

¹⁰Kacperczyk et. al. (2005) and Kacperczyk and Seru (2006) have recently used the same procedure.

¹¹Since fund returns are reported after expenses, we add back annual expenses divided by 12 to reported returns to retrieve monthly before-expense returns.

¹²Data were downloaded from Kenneth French’s website, <http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/>, on June 18, 2004.

¹³Standard errors obtained from a pooled OLS regression with a panel data set are biased downward if residuals are serially correlated within a fund or cross-sectionally within a given period. In addition to including a time dummy, we estimate standard errors clustered by period to correct for these biases (Petersen, 2006). For our sample, standard errors computed in this way are substantially higher than standard errors that do not account for cross-sectional correlation. For example, the standard error reported in Panel A of Table II for the four-factor alpha is 0.3711, while the OLS standard error for the same regression would be 0.1146, the White standard error 0.1236, the standard error clustered by fund 0.1551, and the Newey-West standard error with 12 lags, which corrects for serial correlation, 0.1389. The small difference between the OLS standard error and those adjusted for serial correlation suggests that our approach is adequate for the dataset (see Petersen, 2006). An alternative way to address the contemporaneous correlation in residuals is to employ the Fama-MacBeth procedure (Fama and MacBeth, 1973). This procedure yields standard errors that are close to ours (0.3586, with an estimated coefficient of -1.4792, and 0.2651 with a coefficient of -1.074 if we weight by the number of observations in each month).

¹⁴For the sake of brevity, results are not reported in the paper, but they are available from the authors upon request.

¹⁵The practice of incubation (see Evans, 2006), by which a management company starts several funds with money raised internally and then terminates those funds with low returns and markets those with high returns, may partially explain the omission bias, since management companies “back-fill” the return information of the successful funds, while no information is compiled for the terminated ones.

¹⁶In the original sample of diversified equity mutual funds (including observations with no information on returns or expenses), 15.5% of observations for funds with less than five years in the sample (young funds) do not have data on returns or expenses (the corresponding percentage

for seasoned funds is 5%). Further, the percentage of observations without enough information among young funds becomes larger the earlier the period: if we consider the years 1963-1991, the percentage goes up to 28.9% and it is almost 40% for the period 1963-1980.

¹⁷Sirri and Tufano (1998) are the first to assume a holding period of 7 years. This figure is based on a redemption rate (measured as total outflows in a given year as a percentage of starting assets) of 14% for aggressive growth, long-term growth and growth/income funds in 1990. Data from the Investment Company Institute (2006, page 97) shows that the redemption rate in the 1985-2003 period has been on average 19.5%, implying a holding period of 5.13 years. During the same period, the average redemption rate when outflows include the proceeds reinvested in a fund of the same complex was 37.33%, implying a holding period of 2.67 years.

¹⁸The hypothesis that size and age are negatively related to fund fees has found empirical support in the literature on the determinants of mutual fund fees. See, for instance, Ferris and Chance (1987), Malhotra and McLeod (1997), Luo (2002), or Golec (2003). Recently, however, Chen et al. (2004) have provided evidence that size and performance are negatively related. Their results cast doubt on the potential explanation of our findings as caused by size differences.

¹⁹Henceforth, we focus on Carhart's alpha exclusively. We use α_{it} as a measure of the alpha expected by the manager of fund i at the beginning of period t under the assumption that fees are set at the beginning of period t . If fees were set in the middle of period t , our measure of expected performance would thus aggregate performance observed prior to setting fees with expected performance. We have estimated the fee equation using α_{it+1} as a measure of expected returns and obtained identical results.

²⁰Different aspects of mutual fund fee determination have been studied, among others by Ferris and Chance (1987), Tufano and Sevick (1997), Latzko (1999), Malhotra and McLeod (1997), Luo (2002), Deli (2002), and Golec (2003).

²¹Christoffersen and Musto (2002) do in fact refer to "price sensitivity", rather than "performance sensitivity". For the money market funds that they study, however, they show that fee differences are almost equivalent to net performance differences. In the context of equity funds, fee differences account only partially for differences in net performance.

²²Elton et al. (2001) report a number of errors associated with mutual fund mergers and splits in the CRSP sample. Huang et al. (2006) argue that these errors could lead to extreme values of

flows. We deal with this problem by eliminating the 1% of observations with the lowest and highest flows in each year.

²³In this case, we must first identify share classes of the same fund so a different class is not counted as a different fund.

²⁴We have also estimated equation (14) for holding periods of 2 and 10 years, obtaining similar results.

²⁵We have computed two measures of flow-to-performance sensitivity, each corresponding to a different proxy for participation costs, and then conducted the subsequent analysis with both. Given the similarity of results, we only report those corresponding to complex size as a proxy for participation costs.

²⁶In unreported results, we have found that Q/MAX does not have any significant effect on fund fees if it replaces performance sensitivity in the fee equation, although the associated coefficient is negative.

²⁷Already in 1997, Money Magazine reports: “Over the past decade, the 20% of U.S. diversified stock funds with the highest expenses charged an average of 1.85 % more in annual fees than the cheapest 20%. But those priciest funds earned an annual average of 1.91 % *less than* the cheapest group” (Brush, 1997).

Tables

Table I
Summary Statistics. All funds

The table shows summary statistics for the whole sample of 538,813 fund-month observations. Reported returns are monthly (after-fee) returns in percentage points. *Expenses* denotes funds' expense ratio, and *loads* the maximum of the front-end and back-end load. Both variables are measured in percentage points. *TNA* is total net assets in millions of USD. Values of zero for TNA are due to rounding. Age is in years.

	Mean	Median	S. D.	Max	Min
Returns	0.74	0.99	5.80	50.90	-49.62
Expenses	1.37	1.27	0.61	5.02	0.01
Loads	2.68	1.00	2.91	13.60	0.00
TNA	450.98	53.40	2,428.81	110,525.90	0.00
Age	8.89	5.00	11.62	80.00	0.00

Table II
Before-expense risk adjusted returns and expense ratios.
All funds

The table shows estimated slope coefficients for the OLS regression of funds' risk-adjusted before-expense monthly returns on monthly expense ratios and dummy variables corresponding to each month in the period January 1967-December 2003. The table also reports heteroscedasticity-robust standard errors clustered by time to account for cross-sectional correlation. The total number of fund-month observations is 207,968. One, two and three asterisks indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Performance Measure	Estimated Coefficient	Standard Error	Adjusted R²
Fama-French 3-factor	-0.6257*	0.3710	10.54%
Carhart 4-factor	-1.0101***	0.3715	10.26%

Table III
Before-expense risk adjusted returns and expense ratios.
Actively managed retail mutual funds

The table shows estimated slope coefficients for the OLS regression of funds' risk-adjusted before-expense monthly returns on monthly expense ratios and dummy variables corresponding to each month in the period January 1967-December 2003. The table also reports heteroscedasticity-robust standard errors clustered by time to account for cross-sectional correlation. The total number of fund-month observations is 181,111. One, two and three asterisks indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Performance Measure	Estimated Coefficient	Standard Error	Adjusted R ²
Fama-French 3-factor	-0.6980**	0.3235	10.53%
Carhart 4-factor	-1.0240***	0.3176	10.26%

Table IV
Before-expense risk adjusted returns and expense ratios.
No small funds

The table shows estimated slope coefficients for the OLS regression of funds' risk-adjusted before-expense monthly returns on monthly expense ratios and dummy variables corresponding to each month in the period January 1967-December 2003. The table also reports heteroscedasticity-robust standard errors clustered by time to account for cross-sectional correlation. Fund-month observations with total net assets in the first, second and third deciles of the corresponding month have been removed from the sample. Only actively managed retail funds are included in the sample. One, two and three asterisks indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Performance Measure	Estimated Coefficient	Standard Error	Adjusted R ²	Number of Obs.
First Decile Excluded				
Fama-French 3-factor	-0.3253	0.4002	10.95%	163,505
Carhart 4-factor	-0.9061**	0.3888	10.57%	163,505
First and Second Deciles Excluded				
Fama-French 3-factor	-0.2193	0.4333	10.97%	145,658
Carhart 4-factor	-0.9599**	0.4150	10.59%	145,658
First, Second and Third Decile Excluded				
Fama-French 3-factor	-0.2506	0.4589	10.91%	127,859
Carhart 4-factor	-1.0252**	0.4390	10.54%	127,859

Table V
Before-expense risk adjusted returns and expense ratios.
No-load funds

The table shows estimated slope coefficients for the OLS regression of funds' risk-adjusted before-expense monthly returns on monthly expense ratios and dummy variables corresponding to each month in the period January 1967-December 2003. The table also reports heteroscedasticity-robust standard errors clustered by time to account for cross-sectional correlation. Only actively managed retail funds charging no front-end load or contingent deferred sales load are included in the sample. The total number of fund-month observations is 67,652. One, two and three asterisks indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Performance Measure	Estimated Coefficient	Standard Error	Adjusted R²
Fama-French 3-factor	-0.5128	0.3367	8.84%
Carhart 4-factor	-0.9966***	0.3335	8.66%

Table VI
Before-expense risk adjusted returns and total ownership cost.
Load funds

The table shows estimated slope coefficients for the OLS regression of funds' risk-adjusted before-expense monthly returns on monthly total ownership cost and dummy variables corresponding to each month in the period January 1967-December 2003. Total annual ownership cost is computed by adding annual expense ratios to total loads divided by the holding period. The table also reports heteroscedasticity-robust standard errors clustered by time to account for cross-sectional correlation. Only actively managed retail funds charging a front-end load or a contingent deferred sales load are included in the sample. The total number of fund-month observations is 113,459. One, two and three asterisks indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Performance Measure	Estimated Coefficient	Standard Error	Adjusted R²
Holding period = 2 years			
Fama-French 3-factor	-0.3741***	0.1354	11.77%
Carhart 4-factor	-0.4302***	0.1284	11.42%
Holding period = 7 years			
Fama-French 3-factor	-0.7205*	0.3722	11.77%
Carhart 4-factor	-0.9851***	0.3512	11.43%
Holding period = 10 years			
Fama-French 3-factor	-0.6998*	0.3925	11.77%
Carhart 4-factor	-0.9898***	0.3717	11.43%

Table VII
Before-expense risk adjusted returns and expenses.
Regressions by subperiods

The table shows estimated slope coefficients for the OLS regression of funds' risk-adjusted before-expense monthly returns on annual expenses and dummy variables corresponding to each month in the period January 1967-December 2003. The table also reports heteroscedasticity-robust standard errors clustered by time to account for cross-sectional correlation. Only actively managed retail funds are included in the sample. One, two and three asterisks indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Performance Measure	Estimated Coefficient	Standard Error	Adjusted R²	Number of Obs.
Subperiod 1967-1976				
Fama-French 3-factor	0.5063	0.61590	14.71%	27,202
Carhart 4-factor	-0.4209	0.54970	12.50%	27,202
Subperiod 1977-1985				
Fama-French 3-factor	-0.0716	0.7684	8.23%	21,144
Carhart 4-factor	-1.3395**	0.6570	8.16%	21,144
Subperiod 1986-1994				
Fama-French 3-factor	-0.7167**	0.3268	6.46%	28,832
Carhart 4-factor	-1.0184***	0.3093	5.96%	28,832
Subperiod 1995-2003				
Fama-French 3-factor	-0.9450**	0.4693	10.64%	103,933
Carhart 4-factor	-1.0804**	0.4668	10.68%	103,933

Table VIII
Before-expense risk adjusted returns and expenses.
Regressions by investment objective

The table shows estimated slope coefficients for the OLS regression of funds' risk-adjusted before-expense monthly returns on monthly expenses and dummy variables corresponding to each month in the period January 1992-December 2003. The table also reports heteroscedasticity-robust standard errors clustered by time to account for cross-sectional correlation. Only actively managed retail funds are included in the sample. One, two and three asterisks indicate statistical significance at the 10%, 5% and 1% levels, respectively.

Performance Measure	Estimated Coefficient	Standard Error	Adjusted R²	Number of Obs.
Strategic Insight: Aggressive Growth Funds				
Fama-French 3-factor	-1.8788**	0.8198	18.02%	9,837
Carhart 4-factor	-1.6644**	0.8107	18.79%	9,837
Strategic Insight: Growth MidCap Funds				
Fama-French 3-factor	-1.4858	1.0535	27.64%	9,478
Carhart 4-factor	-0.1146	0.9263	28.62%	9,478
Strategic Insight: Growth and Income Funds				
Fama-French 3-factor	-0.5763***	0.2209	18.00%	29,134
Carhart 4-factor	-0.8118***	0.2343	18.61%	29,134
Strategic Insight: Growth Funds				
Fama-French 3-factor	-1.2230***	0.3584	7.11%	50,349
Carhart 4-factor	-1.0646***	0.3461	7.19%	50,349
Strategic Insight: Small Company Growth Funds				
Fama-French 3-factor	-0.4765	0.5309	28.43%	23,275
Carhart 4-factor	-0.7519	0.5326	24.93%	23,275

Table IX
Fund flows and the flow-to-performance sensitivity

The table shows estimated coefficients for the OLS regression of funds' annual relative growth in assets due to new money on selected fund characteristics in the period January 1992-December 2003. The measure of performance is Carhart's (1997) four-factor alpha in excess of the mean for all funds in the same investment category and year. The table also reports heteroscedasticity-robust standard errors (in parentheses) clustered by time to account for cross-sectional correlation. All fees are expressed in per-unit terms. The total number of fund-year observations is 6,708. One, two and three asterisks indicate statistical significance at the 10%, 5% and 1% levels, respectively.

	Proxy for participation costs	
	Total size of funds in complex	Presence of a star in complex
$size_{t-1}$	-0.0281*** (0.0056)	-0.0283*** (0.0058)
age_{t-1}	-0.0337*** (0.0057)	-0.0339*** (0.0059)
$FLoad_{t-1}$	-0.1290 (0.5491)	-0.1420 (0.5629)
$BLoad_{t-1}$	0.1009 (0.8189)	0.0852 (0.8058)
$Mfee_{t-1}$	-0.3856 (1.6876)	-0.5338 (1.5829)
$12b1_{t-1}$	-4.7736** (2.0348)	-4.6017** (2.0444)
$classA_{t-1}$	0.0270 (0.0280)	0.0279 (0.0308)
$classB_{t-1}$	-0.0740 (0.0534)	-0.0738 (0.0529)
$classC_{t-1}$	0.0621* (0.0363)	0.0623* (0.0371)
σ_{t-1}	-0.3154 (0.2308)	-0.3042 (0.2160)

Table IX– *Continued*

	Total size of funds in complex	Presence of a star in complex
$size_complex_{t-1}$	0.0140*** (0.0042)	0.0126** (0.0050)
$flow_{t-1}$	0.0837** (0.0341)	0.0835** (0.0340)
$objflow_{t-1}$	0.3505*** (0.1271)	0.3516*** (0.1238)
Q/Max_{t-1}	0.2586*** (0.0594)	0.2546*** (0.0596)
$Perf_{t-1}$	0.7404** (0.2907)	0.7567** (0.3050)
$Perf_{t-1} \cdot rel_age_{t-1}$	-0.2268** (0.1152)	-0.2601* (0.1356)
$Perf_{t-1} \cdot rel_Q/MAX_{t-1}$	1.0516*** (0.2051)	0.9012*** (0.2425)
$I_{H,t-1}$	0.0650*** (0.0250)	0.0528* (0.0299)
$Perf_{t-1} \cdot I_{H,t-1}$	0.7873*** (0.1641)	1.3202*** (0.3166)
$PC_{t-1} \cdot I_{H,t-1}$	0.0023 (0.0062)	0.0211 (0.0221)
$Perf_{t-1} \cdot PC_{t-1} \cdot I_{H,t-1}$	-0.1339*** (0.0441)	-0.8269*** (0.3070)
$Adjusted R^2$	12.42%	12.36%

Table X
Determinants of total ownership cost and management fees

The table shows estimated coefficients for the OLS regression of funds' fees on selected fund characteristics in the period January 1992-December 2003. In columns (1)-(2), the dependent variable is total annual ownership cost. This cost is computed for all funds, except those of class B or C, by adding total loads divided by a holding period of 7 years to annual expense ratios. For class-B or -C funds, back-end loads are subtracted from total loads. The dependent variable in column (3) is management fees defined as the difference between the expense ratio and the 12b-1 fee. All fees are expressed in bp. The table also reports heteroscedasticity-robust standard errors (in parentheses) clustered by time to account for cross-sectional correlation. $\hat{\alpha}_t$ is the year t 's 4-factor alpha. The total number of fund-year observations is 8829, 6675, and 6378 for columns (1), (2), and (3), respectively. One, two and three asterisks indicate statistical significance at the 10%, 5% and 1% levels, respectively.

	(1)	(2)	(3)
$size_{t-1}$	-9.5024*** (1.1888)	-13.4112*** (2.7120)	-7.4601*** (2.0775)
$size_{t-1}^2$	0.2293** (0.1061)	0.58498*** (0.2129)	0.2937** (0.1480)
age_{t-1}	-3.3418*** (0.9238)	-6.9389*** (0.6780)	-5.8075*** (.72267)
$size_complex_{t-1}$	-5.3533*** (1.1635)	-4.5914*** (1.2498)	-13.4956*** (0.4742)
$size_complex_{t-1}^2$	0.15723** (0.0662)	0.07020 (.08421)	0.4751*** (0.03397)
$n_complex_{t-1}$	-0.08226*** (0.0308)	-0.14013*** (.03878)	0.09984*** (0.3104)
$turn_{t-1}$	3.9733*** (0.4299)	4.2772*** (.70323)	3.7286*** (0.7030)
σ_{t-1}	35.6745*** (12.4637)	30.4614*** (8.6865)	12.1905*** (8.0574)

Table X– *Continued*

GMC_{t-1}	-10.1372*** (1.3860)	-8.6189*** (1.6625)	-11.9669*** (0.9400)
GRI_{t-1}	-16.6996*** (1.2606)	-14.9580*** (1.2300)	-19.4193*** (0.85473)
GRO_{t-1}	-9.8955*** (1.0104)	-8.9809*** (1.4889)	-11.7772*** (1.1546)
SCG_{t-1}	-4.0166*** (1.0274)	-3.7789** (1.6308)	-4.3689*** (0.8291)
$Load_single - class_{t-1}$	74.8849*** (1.6098)	74.7276*** (2.1866)	-1.6398 (1.4505)
$classA_{t-1}$	92.2276*** (1.1110)	92.8206*** (1.2121)	4.1427*** (0.6261)
$classB_{t-1}$	85.9762*** (1.7369)	84.3982*** (2.1803)	3.4888 (2.2649)
$classC_{t-1}$	80.1799*** (3.2813)	78.8592*** (4.0051)	0.6761 (3.2667)
$\hat{\alpha}_t$	-9.9144** (4.6003)	-12.7013*** (4.1554)	-13.0515*** (4.0220)
\hat{S}_{t-1}		-6.7943*** (1.10640)	-6.45312*** (1.0538)
$Adjusted R^2$	65.17%	67.54%	43.84%

Figure 1
Mutual fund expenses and loads

The figure shows yearly averages of mutual fund expenses (solid line) and loads (dashed line) for the 1962-2003 period. It also displays yearly averages of loads among funds charging positive loads (dotted line).

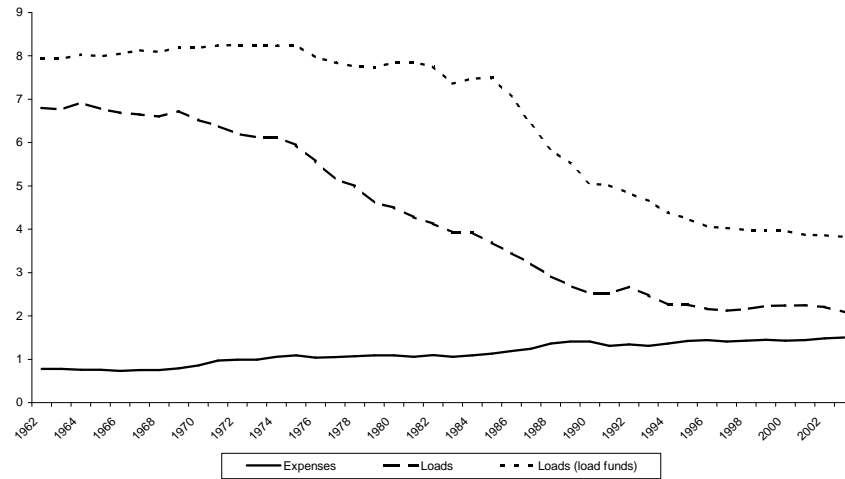


Figure 2
Total Net Assets

The figure shows yearly averages of the total net assets (measured in billions of USD.) managed by the mutual funds in the sample.

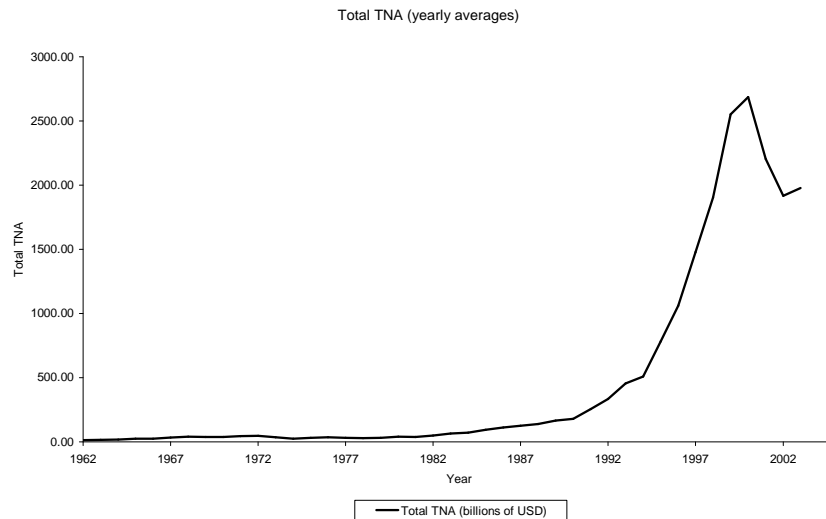


Figure 3
Number of funds

The figure shows yearly averages of the number of funds in the sample. The dashed line counts each fund-class as an individual fund (as reported in the dataset). The solid line aggregates the different classes of a single fund into one fund, and counts the number of these funds. The two lines coincide until year 1991.

